

From ethical AI frameworks to tools: a review of approaches

Year: 2023 | Citations: 134 | Authors: E. Prem

Abstract

In reaction to concerns about a broad range of potential ethical issues, dozens of proposals for addressing ethical aspects of artificial intelligence (AI) have been published. However, many of them are too abstract for being easily translated into concrete designs for AI systems. The various proposed ethical frameworks can be considered an instance of principlism that is similar to that found in medical ethics. Given their general nature, principles do not say how they should be applied in a particular context. Hence, a broad range of approaches, methods, and tools have been proposed for addressing ethical concerns of AI systems. This paper presents a systematic analysis of more than 100 frameworks, process models, and proposed remedies and tools for helping to make the necessary shift from principles to implementation, expanding on the work of Morley and colleagues. This analysis confirms a strong focus of proposed approaches on only a few ethical issues such as explicability, fairness, privacy, and accountability. These issues are often addressed with proposals for software and algorithms. Other, more general ethical issues are mainly addressed with conceptual frameworks, guidelines, or process models. This paper develops a structured list and definitions of approaches, presents a refined segmentation of the AI development process, and suggests areas that will require more attention from researchers and developers.